Appendices for "Keep Calm and Switch On! Preserving Sentiment and Fluency in Semantic Text Exchange"

Steven Y. Feng* **Aaron W. Li*** **Jesse Hoey**

David R. Cheriton School of Computer Science
University of Waterloo
Waterloo, Ontario, Canada
{sy2feng, w89li, jhoey}@uwaterloo.ca

A Masked Output Examples

Table[1] 1 includes various example masked outputs by the ERM and SMM modules. We illustrate a variety of word-to-word, phrase-to-word, word-to-phrase, and phrase-to-phrase entity replacements and similarity masking.

B Dataset Statistics

See Table 2 for our training and testing splits by dataset and sentiment.

C Details of Baseline Implementations

C.1 NWN-STEM

This follows Algorithm 2 in Yao et al. (2017). For each dataset and each RE (note that this model is restricted to noun REs), we go through the dataset's training set (which acts as the "reference review set") and extract a list of text lines that contain the RE (where the RE acts as the "topic keyword"). For each of these lines, with the help of the Stanford Parser, we extract all single-word nouns, and for each of them (which we call $noun_i$), we check if $MIN_{sim}(noun_i, RE) > 0.1$. If so, we add them to the list of nouns similar to the RE, which we call sim_{nouns}.

For each evaluation line and associated RE, we extract all singular nouns within the text that are similar to the RE. For our evaluation purposes, we choose two MIN_{sim} values of 0.075 and 0 to produce two outputs per input. These two MIN_{sim} values result in actual replacement rates similar to the actual masking/replacement rates of other models (SMERTI and W2V-STEM) for MRT/RRT of 20% and 40%, respectively. Each similar noun is replaced with the noun in sim_{nouns} that is most similar to it to produce the output text.

When analyzing every word during the above procedure, we take the first default WordNet synset's definition of the word (e.g. "bill" will default to "a statute in draft before it becomes law"). Further, if a word does not exist in WordNet, we use WordNet's Morphy to find if a lexically similar word exists that may just be in a different form (e.g. "photos" does not exist in WordNet, but using Morphy, we can find "photograph") and use the first default synset of this resulting word.

Note that since the number of nouns per input is limited, the actual replacement rates have an upper limit, and this is why we only generate two outputs per input to compare to outputs of other models for 20% and 40% MRT/RRT.

C.2 GWN-STEM

This is a modification of NWN-STEM to handle verbs and adjectives (note that WordNet only works for single words). On top of nouns, we also extract similar verbs and adjectives to the RE. This results in three lists: sim_{nouns}, sim_{verbs}, and sim_{adjs}, where we still use a MIN_{sim} value of 0.1 for determining the three lists. Further, we not only replace nouns in the original text, but also verbs and adjectives. Note that we use the same synset and Morphy procedure as for NWN-STEM. For our evaluation, we choose the same MIN_{sim} values of 0.075 and 0 for noun REs, but MIN_{sim} values of 0.1 and 0 for verb REs. These combinations of MIN_{sim} values result in actual replacement rates similar to the actual masking/replacement rates of other models for MRT/RRT of 20% and 40%, respectively.

We noticed that GWN-STEM only works for noun and verb REs, as for most adjectives, WordNet cannot calculate similarity scores. Hence, it was infeasible to evaluate on adjective REs. Further, most similarity scores only exist between noun-noun pairs and verb-verb pairs, and when we

*Authors contributed equally

[1] Tables and figures mentioned in this appendices document refer to the tables and figures here

S	*family enjoyed the food quite a bit especially the sweet and sour chicken .*
RE	*bitter*
S''_1	*family enjoyed the food quite a bit especially the sweet and bitter chicken .*
S''_2	*family [mask] the [mask] a bit especially the sweet and bitter [mask] .*
S''_3	*[mask] the [mask] bit [mask] the [mask] and bitter [mask] .*
S''_4	*[mask] bit [mask] bitter [mask] .*
S	*terrible customer service . couldn't make a wire transfer because they are out of paper .*
RE	*amazing*
S''_1	*amazing customer [mask] . couldn't make a wire transfer [mask] they are out of paper .*
S''_2	*amazing customer [mask] . couldn't make a wire transfer [mask] they [mask] out of paper .*
S''_3	*amazing [mask] . [mask] make a wire transfer [mask] out [mask] paper .*
S''_4	*amazing [mask] . [mask] a [mask] out [mask] .*
S	*Heather enjoyed her movie date with Jim last night.*
RE	*yesterday*
S''_1	*heather [mask] her movie [mask] with jim yesterday .*
S''_2	*heather [mask] her movie [mask] with jim yesterday .*
S''_3	*heather [mask] her movie [mask] with jim yesterday .*
S''_4	*[mask] yesterday .*
S	*Heather enjoyed her movie date with Jim last night.*
RE	*Elizabeth*
S''_1	*elizabeth enjoyed [mask] movie date with [mask] last night .*
S''_2	*elizabeth enjoyed [mask] movie date with [mask] .*
S''_3	*elizabeth enjoyed [mask] movie date with [mask] .*
S''_4	*elizabeth enjoyed [mask] with [mask] .*
S	*The car crashed into the building and exploded, killing hundreds.*
RE	*caught on fire*
S''_1	*the car [mask] caught on fire and [mask] , killing hundreds .*
S''_2	*[mask] caught on fire and [mask] , killing hundreds .*
S''_3	*[mask] caught on fire and [mask] , killing hundreds .*
S''_4	*[mask] caught on fire [mask] .*
S	*My son took his math test yesterday and failed. He cried all day and I hate him now.*
RE	*medical examination*
S''_1	*[mask] took medical examination yesterday and [mask] . he cried all day and i hate [mask] now .*
S''_2	*[mask] took medical examination yesterday and [mask] . he cried all day and i hate [mask] now .*
S''_3	*[mask] took medical examination yesterday and [mask] . he cried all day and i hate [mask] now .*
S''_4	*[mask] medical examination [mask] and [mask] . [mask] .*
S	*I took my dog for a walk in the park. He really enjoyed it!*
RE	*the river*
S''_1	*i took my dog for [mask] in the river . he really enjoyed it !*
S''_2	*i took [mask] for [mask] in the river . he really enjoyed it !*
S''_3	*i [mask] for [mask] in the river . he [mask] !*
S''_4	*i [mask] the river . he [mask] !*
S	*It is very sunny outside so I am very sweaty.*
RE	*extremely snowy*
S''_1	*it is extremely snowy outside so i am [mask] .*
S''_2	*it is extremely snowy [mask] so i am [mask] .*
S''_3	*it [mask] extremely snowy [mask] i [mask] .*
S''_4	*[mask] extremely snowy [mask] .*
S	*I went to my friend Amy's house last night.*
RE	*my husband Jim's*
S''_1	*i went to my husband jim's house last night .*
S''_2	*i went to my husband jim's house [mask] .*
S''_3	*i went to my husband jim's [mask] .*
S''_4	*i [mask] my husband jim's [mask] .*
S	*I went to my friend Amy's house last night.*
RE	*my husband Jim's boat*
S''_1	*i went to my husband jim's boat last night .*
S''_2	*i went to my husband jim's boat last night .*
S''_3	*i went to my husband jim's boat [mask] .*
S''_4	*i [mask] my husband jim's boat [mask] .*

Table 1: Example masked outputs. S is the original input text; RE is the replacement entity; S''_1 corresponds to $MRT = 0.2$, base $ST = 0.4$; S''_2 corresponds to $MRT = 0.4$, base $ST = 0.3$; S''_3 corresponds to $MRT = 0.6$, base $ST = 0.2$; S''_4 corresponds to $MRT = 0.8$, base $ST = 0.1$

Dataset	Sentiment	Training Set	Testing Set
Amazon	Positive	30K	5K
	Negative	30K	5K
	Neutral	15K	2.5K
Yelp	Positive	30K	5K
	Negative	30K	5K
	Neutral	15K	2.5K
News Headlines	-	120K	20K

Table 2: Training and testing splits by dataset

tried to produce sim_{verbs} and sim_{adjs} for noun REs, almost all resulted in empty lists. Hence, GWN-STEM actually produced the same outputs as NWN-STEM for noun REs (and was also limited to two replacement rates), which is why the two models have the same outputs and resulting metrics for noun REs. Even for verb REs, we were limited to two sets of outputs (mimicking the two replacement rates above) since similarity calculations between verb-noun pairs and verb-adjective pairs were limited, so few were replaced.

C.3 W2V-STEM

This uses Word2Vec (W2V) models trained using Gensim. We train six W2V models: one unigram model per dataset, and one four-gram model per dataset, where each is trained using the corresponding dataset's training set. To train the four-gram models, we begin by applying a bi-gram phrasing model on top of the original text, and then the bi-gram phrasing model again on top of this resulting text. We call this a four-gram phrasing model. We then use this to generate text that is grouped into phrases up to four-grams long. We then train W2V models on this four-gram text to generate the four-gram W2V models.

For the unigram models, we use an embedding vector size of 50, a context window of 3, a minimum token count of 0, and the skip-gram model. For the four-gram models, we use an embedding vector size of 10, a context window of 1, a minimum token count of 0, and the CBOW (continuous bag-of-words) model.

For evaluation lines with noun, verb, and adjective REs, we go through the line of text and with the help of the Stanford parser, extract all words that are the same POS as the RE (which become the candidate OEs). Then, using cosine similarity between the W2V embedding vectors (from the unigram W2V models) of the RE and candidate OEs, we find the word with the maximum similarity to the RE which becomes the replaced OE. Then, we replace other words in the input text similar to the OE using vector addition and subtraction as described in Section 4.3. We start with a similarity threshold value of 0.1 that increases by 0.05 each time to generate text satisfying varying replacement rate thresholds.

For evaluation lines with phrase REs, we first generate text files of the evaluation sets that are grouped into phrases up to four-grams long using the four-gram phrasing model. Then, using the four-gram W2V models, we proceed similarly as above to determine the replaced OE, which can now be either a single word or a phrase. Other words and phrases in the input text are replaced using vector addition and subtraction. We start with a similarity threshold value of 0.3 that increases by 0.01 each time to generate text satisfying varying replacement rate thresholds.

D Evaluation Keywords/Phrases

We select ten keywords (five keyphrases) to act as our REs per dataset per POS. See Tables 3 to 6 for the chosen REs. To do so, we iterate through our test sets and with the Stanford Parser, we extract a list of nouns and noun phrases, verbs and verb phrases, and adjectives and adjective phrases. We sort these lists by frequency, and limit our selections to the top 10% most frequent. For the verbs and adjectives and their phrases, we further filter them through a list of sentiment words (Hu and Liu, 2004) to ensure the REs we choose do not carry significant sentiment-related meaning, as inserting them into the original text would obviously lead to major changes in sentiment. From these, we manually select ten per dataset per POS (except phrases, where we select five per dataset) that are significant and carry strong meaning. These work well as the REs for evaluation purposes. Note that for phrases, we choose three noun phrases, one verb phrase, and one adjective phrase per dataset.

We choose from the most frequent words and phrases as they are more common and likely hold more significant meaning compared to less frequent ones (e.g. names and typos). Manual selection was required as some of the most frequent words/phrases hold little semantic meaning (e.g. *it*, *they*, *is*, *was*, and so forth). We only choose half the number of phrases as words as we find that frequent phrases carrying significant semantic meaning with little sentiment are much rarer.

Yelp	Amazon	News Headlines
Food	Book	Trump
Service	Product	Photos
Place	Time	Video
Staff	Price	World
Time	Quality	Women
Customer	Money	Life
Atmosphere	Game	Kids
Pizza	Story	People
Restaurant	Movie	Week
Chicken	Phone	Obama / President*

Table 3: Chosen evaluation noun REs. **Obama* does not exist in WordNet, so we instead use the word *President* for NWN-STEM and GWN-STEM.

Yelp	Amazon	News Headlines
Ordered	Buy	Live
Closed	Ordered	Think
Tasted	Received	Make
Return	Expected	Stop
Waiting	Purchased	Watch
Serve	Reading	Save
Eating	Advertised	Avoid
Visiting	Install	Learn
Looking	Playing	Create
Priced	Recommend	Teach

Table 4: Chosen evaluation verb REs

Yelp	Amazon	News Headlines
Small	Different	Black
Little	Light	White
New	Predictable	Old
Big	Little	American
Mexican	Heavy	National
First	Thick	Sexual
Long	Plastic	Single
High	Old	Global
Open	New	Presidential
Busy	First	Female

Table 5: Chosen evaluation adjective REs

Yelp	Amazon	News Headlines
Customer Service	My Daughter	Donald Trump
Your Money	Another One	Hillary Clinton
Las Vegas	This Game	Climate Change
Come Back	Put it Down	Need to Know
Very Small	Much Smaller	Most Important

Table 6: Chosen evaluation phrase REs

E Detailed Evaluation Results - Tables and Graphs

See Figure 1 for a graph of overall average results, Table 7 and Figure 2 for average results by POS, Table 8 and Figure 3 for average results by dataset, and Table 9 and Figure 4 for average results by MRT/RRT. Note that the bolded values in the tables show which model performs better on that particular metric, on average, for the category.

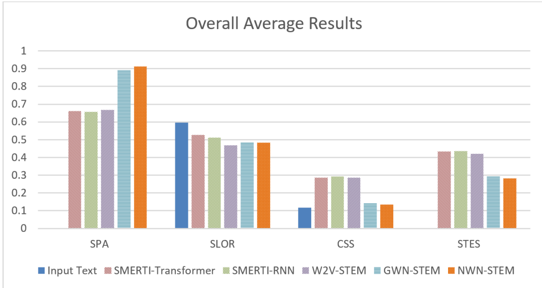

Figure 1: Graph of overall average results (referring to the data found in Table 2 of the main body)

F Model Output Examples

See Tables 10 to 21 for example outputs from every model for all datasets and POS.

References

Minqing Hu and Bing Liu. 2004. Mining and summarizing customer reviews. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 168–177. ACM.

Yuanshun Yao, Bimal Viswanath, Jenna Cryan, Haitao Zheng, and Ben Y Zhao. 2017. Automated crowdturfing attacks and defenses in online review systems. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*, pages 1143–1158. ACM.

		Nouns			Verbs			Adjectives			Phrases		
Model	**MRT/ RRT**	**SPA**	**SLOR**	**CSS**	**SPA**	**SLOR**	**CSS**	**SPA**	**SLOR**	**CSS**	**SPA**	**SLOR**	**CSS**
Input	---	---	0.5971	0.1175	---	0.5960	0.1287	---	0.5977	0.0935	---	0.5939	0.1267
SMERTI-Transformer	20%	0.8560	0.5493	0.2591	0.7887	0.5358	0.2152	0.7517	0.5421	0.2003	0.7971	0.5806	0.3028
	40%	0.7610	0.5382	0.2683	0.7143	0.5286	0.2215	0.6688	0.5326	0.2059	0.7060	0.5753	0.3178
	60%	0.6707	0.5200	0.2899	0.6328	0.5193	0.2461	0.5798	0.5091	0.2286	0.6016	0.5661	0.3516
	80%	0.5638	0.4592	0.3613	0.5003	0.4783	0.3409	0.4832	0.4360	0.3222	0.4933	0.5379	0.4395
	AVG	0.7129	**0.5167**	0.2947	0.6590	**0.5155**	0.2559	**0.6209**	**0.5049**	0.2392	0.6495	**0.5650**	0.3529
	STES		0.4456			**0.4073**			0.3861			0.4884	
SMERTI-RNN	20%	0.8568	0.5443	0.2620	0.7946	0.5299	0.2170	0.7576	0.5338	0.2048	0.7924	0.5751	0.3075
	40%	0.7602	0.5273	0.2725	0.7204	0.5145	0.2243	0.6641	0.5161	0.2109	0.7040	0.5673	0.3242
	60%	0.6688	0.5023	0.2979	0.6290	0.4967	0.2507	0.5792	0.4871	0.2365	0.5996	0.5565	0.3604
	80%	0.5570	0.4397	0.3744	0.4889	0.4590	0.3477	0.4718	0.4139	0.3376	0.4736	0.5313	0.4541
	AVG	0.7107	0.5034	0.3017	0.6582	0.5000	0.2599	0.6182	0.4877	0.2474	0.6424	0.5576	**0.3615**
	STES		**0.4472**			**0.4073**			**0.3891**			**0.4905**	
W2V-STEM	20%	0.9038	0.5342	0.2804	0.8207	0.5140	0.2360	0.7724	0.5158	0.2251	0.8987	0.6078	0.2992
	40%	0.7628	0.4734	0.3063	0.7160	0.4484	0.2694	0.6348	0.4525	0.2490	0.7693	0.5626	0.2954
	60%	0.6354	0.4216	0.3232	0.6124	0.3941	0.2965	0.5108	0.4016	0.2668	0.6280	0.5237	0.2913
	80%	0.5233	0.3864	0.3374	0.5168	0.3551	0.3170	0.4209	0.3702	0.2794	0.5411	0.5136	0.2892
	AVG	0.7063	0.4539	**0.3118**	0.6665	0.4279	**0.2797**	0.5847	0.4350	**0.2551**	**0.7093**	0.5519	0.2938
	STES		0.4395			0.4047			0.3784			0.4528	
GWN-STEM	20%	0.9291	0.5043	0.1288	0.9177	0.5421	0.1372						
	40%	0.8941	0.4621	0.1381	0.8203	0.4372	0.1635						
	AVG	**0.9116**	0.4832	0.1335	**0.8690**	0.4896	0.1504						
	STES		0.2814			0.3048							
NWN-STEM	20%	0.9291	0.5043	0.1288									
	40%	0.8941	0.4621	0.1381									
	AVG	**0.9116**	0.4832	0.1335									
	STES		0.2814										

Table 7: Average results by POS

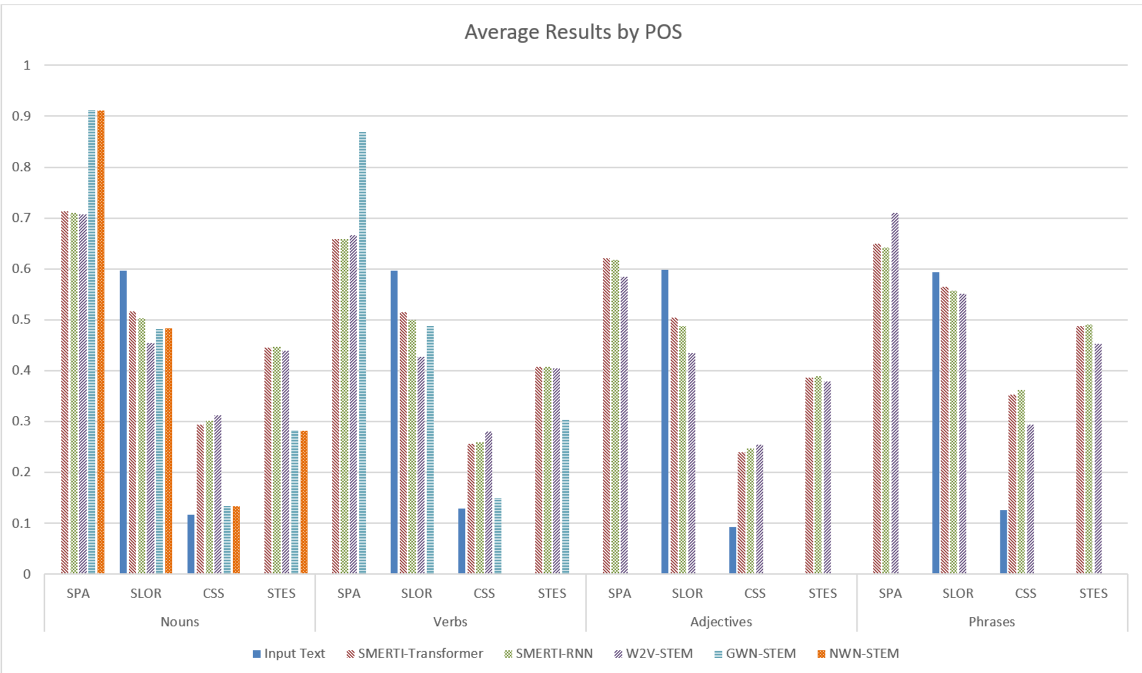

Figure 2: Graph of average results by POS

Model	MRT/RRT	Amazon Reviews			Yelp Reviews			News Headlines		
		SPA	SLOR	CSS	SPA	SLOR	CSS	SPA	SLOR	CSS
Input	---	---	0.6230	0.1113	---	0.5532	0.1429	---	0.6123	0.0955
SMERTI-Transformer	20%	0.8140	0.5773	0.2336	0.8544	0.5162	0.2497	0.7267	0.5623	0.2497
	40%	0.7168	0.5709	0.2410	0.7800	0.5055	0.2577	0.6408	0.5547	0.2615
	60%	0.6363	0.5550	0.2663	0.6981	0.4804	0.2759	0.5293	0.5503	0.2950
	80%	0.5408	0.4776	0.3598	0.5781	0.4045	0.3370	0.4116	0.5514	0.4011
	AVG	0.6770	**0.5452**	0.2752	0.7276	**0.4767**	0.2801	0.5771	**0.5547**	0.3018
	STES	0.4319			0.4260			0.4380		
SMERTI-RNN	20%	0.8113	0.5779	0.2354	0.8488	0.5174	0.2522	0.7409	0.5420	0.2559
	40%	0.7198	0.5634	0.2441	0.7733	0.5037	0.2621	0.6434	0.5268	0.2677
	60%	0.6318	0.5390	0.2714	0.6970	0.4786	0.2831	0.5286	0.5144	0.3047
	80%	0.5195	0.4636	0.3709	0.5608	0.4058	0.3505	0.4132	0.5136	0.4140
	AVG	0.6706	0.5360	**0.2804**	0.7200	0.4764	0.2870	0.5815	0.5242	**0.3106**
	STES	**0.4333**			**0.4302**			**0.4382**		
W2V-STEM	20%	0.8628	0.5649	0.2526	0.8868	0.5158	0.2661	0.7971	0.5481	0.2619
	40%	0.7468	0.4902	0.2728	0.7675	0.4634	0.2865	0.6478	0.4990	0.2808
	60%	0.6201	0.4258	0.2867	0.6391	0.4167	0.3007	0.5308	0.4633	0.2960
	80%	0.5185	0.3898	0.2978	0.5316	0.3879	0.3118	0.4515	0.4413	0.3077
	AVG	0.6870	0.4677	0.2775	0.7063	0.4459	**0.2912**	0.6068	0.4880	0.2866
	STES	0.4168			0.4230			0.4175		
GWN-STEM	20%	0.9605	0.5509	0.1232	0.9638	0.4924	0.1674	0.8458	0.5262	0.1085
	40%	0.9078	0.4649	0.1330	0.9347	0.4250	0.1783	0.7292	0.4591	0.1410
	AVG	0.9342	0.5079	0.1281	0.9493	0.4587	0.1729	0.7875	0.4926	0.1248
	STES	0.2766			0.3327			0.2651		
NWN-STEM	20%	0.9573	0.5287	0.1042	0.9693	0.4732	0.1645	0.8607	0.5110	0.1178
	40%	0.9337	0.4818	0.1182	0.9537	0.4302	0.1727	0.7950	0.4741	0.1234
	AVG	**0.9455**	0.5053	0.1112	**0.9615**	0.4517	0.1686	**0.8278**	0.4926	0.1206
	STES	0.2493			0.3266			0.2602		

Table 8: Average results by dataset

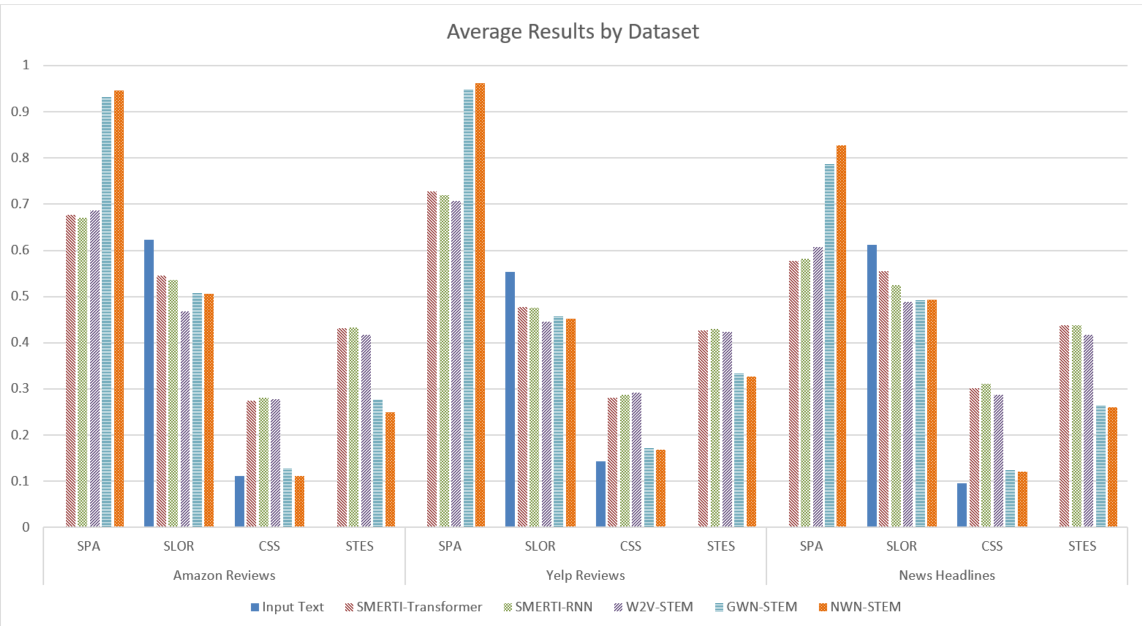

Figure 3: Graph of average results by dataset

		Overall Averages				
Model	**MRT/ RRT**	**Actual MR/RR**	**SPA**	**SLOR**	**CSS**	**STES**
Input	---	---	---	0.5962	0.1166	---
SMERTI-Transformer	20%	13.98%	0.7984	0.5519	0.2443	0.4192
	40%	31.58%	0.7125	0.5437	0.2534	0.4173
	60%	51.50%	0.6212	0.5286	0.2791	0.4234
	80%	74.58%	0.5102	0.4778	0.3659	0.4421
	AVG	42.91%	0.6606	**0.5255**	0.2857	0.4337
SMERTI-RNN	20%	13.99%	0.8003	0.5458	0.2478	0.4215
	40%	31.61%	0.7122	0.5313	0.2580	0.4189
	60%	51.55%	0.6191	0.5107	0.2864	0.4246
	80%	74.60%	0.4978	0.4610	0.3785	0.4399
	AVG	42.94%	0.6574	0.5122	**0.2927**	**0.4354**
W2V-STEM	20%	12.98%	0.8489	0.5429	0.2602	0.4371
	40%	29.00%	0.7207	0.4842	0.2800	0.4271
	60%	48.95%	0.5967	0.4353	0.2945	0.4071
	80%	70.01%	0.5005	0.4063	0.3057	0.3881
	AVG	40.24%	0.6667	0.4672	0.2851	0.4197
GWN-STEM	20%	16.52%	0.9234	0.5232	0.1330	0.2854
	40%	33.85%	0.8572	0.4496	0.1508	0.2993
	AVG	25.18%	0.8903	0.4864	0.1419	0.2934
NWN-STEM	20%	19.05%	0.9291	0.5043	0.1288	0.2772
	40%	30.26%	0.8941	0.4621	0.1381	0.2851
	AVG	24.65%	**0.9116**	0.4832	0.1335	0.2814

Table 9: Average results by MRT/RRT

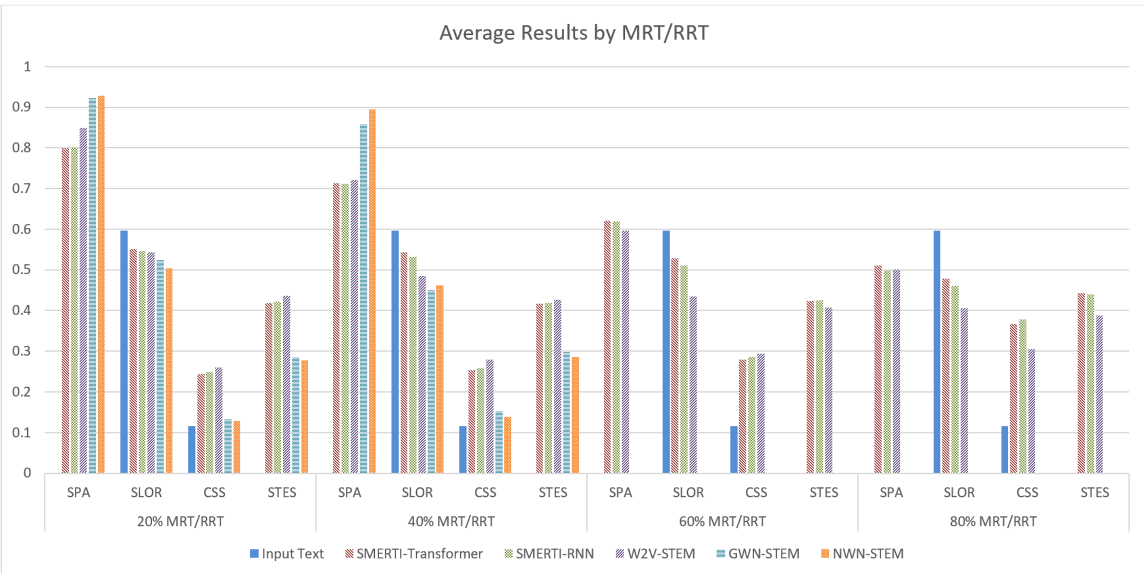

Figure 4: Graph of average results by MRT/RRT

Amazon - Noun *RE*	
Input:	bought this in a few colors and simply love the style . highly recommend purchasing this as well as other colors .
RE:	book
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	bought this in a few chapters and i love the book . highly recommend purchasing this as well as the product .
40%	bought this in my kindle and absolutely love the book . highly recommend purchasing this as a gift as well .
60%	bought this in my kindle and absolutely love the book . highly recommend purchasing this as a gift as well .
80%	bought this book in august and love the book . i highly recommend purchasing it .
SMERTI-RNN	
20%	bought this in a few months and i love the book . highly recommend purchasing this as well as needed .
40%	bought this in my style and i love the book . highly recommend purchasing this as a gift as advertised .
60%	bought this in my review and went through in a leg and in the book . i recommend purchasing this .
80%	bought this in my review and went through in a leg and in the book . i recommend purchasing this .
W2V-STEM	
20%	bought this in a few item and simply love the book . highly recommend purchasing this as well as other item.
40%	bought this in a few item application thing episode the book . highly recommend purchasing this as well as other item .
60%	bought preview page a few item application thing episode page book . despite reccomend purchasing preview as fully as other item .
80%	purchased preview page information few item application thing episode page book . despite reccomend refunded preview possible fully possible reviews item .
GWN-STEM / NWN-STEM	
20% (0.075 MIN_{sim})	bought this in a few decal and simply love the style . highly recommend purchasing this as well as other decal .
40% (0 MIN_{sim})	bought movie in a few decal and simply love the locations . highly recommend purchasing movie as well as other decal .

Table 10: Example outputs for an Amazon evaluation line with noun RE

Amazon - Verb *RE*	
Input:	i didnt even get these . and i think i ordered them like last year . and i think that is ridiculous !
RE:	buy
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	i don't even buy these . and i think i ordered a like last year . and i think that was ridiculous !
40%	do not buy these . and i ordered the wrong one for my year . and they are ridiculous !
60%	do not buy this product . i ordered another pair . it's ridiculous !
80%	don't buy this product . terrible screen protector . will not buy again !
SMERTI-RNN	
20%	i don't even buy these . and i think i ordered more like last year . and i think that it's ridiculous !
40%	i would buy these . and i ordered them like a year . and from a ridiculous !
60%	don't buy this product . i ordered a x . w extremely ridiculous !
80%	don't buy this . a good rip . recommended !
W2V-STEM	
20%	i didnt even buy these . and i yes i ordered them like last year . and i yes that is ridiculous !
40%	purchasing didnt buyers buy these . definitely purchasing yes purchasing ordered them like last year . definitely purchasing yes that smells ridiculous !
60%	purchasing didn't buyers buy these . definitely purchasing yes purchasing bought theses like chinese year . definitely purchasing yes yes smells worthy !
80%	purchasing didn't buyers buy pair . definitely purchasing yes purchasing bought theses recomend chinese yrs . definitely purchasing yes yes smells worthy !
GWN-STEM	
20% (0.1 MIN_{sim})	i didnt even get these . and i think i ordered them like last yrs . and i think that was ridiculous !
40% (0 MIN_{sim})	i didnt even get knowledge . and i think i requested knowledge like last yrs . and i think knowledge was ridiculous !

Table 11: Example outputs for an Amazon evaluation line with verb RE

Amazon - Adjective *RE*	
Input:	was way too small . i had ordered a xlt , it probably was a medium .
RE:	predictable
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	was not too predictable . i had ordered a wow , it probably was a waste of money .
40%	was too predictable . i had ordered a wow , but the book probably was a waste of time .
60%	was too predictable . i had ordered a wow , but the book probably was a waste of time .
80%	easily predictable . i like a twist , it took a long time to finish .
SMERTI-RNN	
20%	was too predictable . i had ordered a refund , it probably was a little disappointing .
40%	was too very predictable . i had ordered a refund , but probably was a disappointing format .
60%	was too very predictable . i had ordered a refund , but probably was a disappointing format .
80%	very predictable . i like a refund , but a little heavy .
W2V-STEM	
20%	was way too predictable . i had ordered a xlt , it probably was a plot .
40%	was plot plot predictable . i had ordered a xlt , it probably was a plot .
60%	was plot plot predictable . i had trilogy action xlt , plot author was action plot .
80%	ending plot plot predictable . plot had trilogy action xlt , plot author ending action plot .

Table 12: Example outputs for an Amazon evaluation line with adjective RE

Amazon - Phrase *RE*	
Input:	bought for my friendz's baby i hope she likes it . it looks cute . baby will like it . small small cute
RE:	my daughter
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	bought for my daughter i hope she likes it . it looks cute . she will like it . very nice
40%	bought for my daughter i think she likes it . it fits great . she will like it . don't recommend
60%	i bought this for my daughter i and she loved it . it fits great too . i will like it . thanks
80%	i bought this for my daughter for christmas and she loves it . she like it . the color is great
SMERTI-RNN	
20%	bought for my daughter i hope she likes it . it looks cute . i will like it . thank you
40%	bought for my daughter i and she likes it . it worked well . i will like it . thank you
60%	bought this item for my daughter and she loves it . she loves it . she will like it again . i love it
80%	love my daughter loves it . very good . he
W2V-STEM	
20%	bought for my daughter friendz's baby i hope she likes it it looks cute baby will like it small small cute
40%	bought for my daughter friendz's thing i hope she likes great smell cute baby will like great quality quality if
60%	bought for my daughter friendz's thing i would not rock solid great smell cute baby will like great quality quality if
80%	balinese my daughter friendz's thing i would not rock solid great smell cute baby as easy great quality quality if

Table 13: Example outputs for an Amazon evaluation line with phrase RE

Yelp - Noun *RE*	
Input:	i love this place! very nice people running the cafe and the food is always good . stars!
RE:	service
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	i love this place! very nice people running the service and the food is always good. stars!
40%	i love this place! very nice people running the service and the food is always good. food is delicious!
60%	i love this place! very nice people running the service and the food is always good. food is delicious!
80%	i love this place! great service and food. best egg rolls!
SMERTI-RNN	
20%	i love this place! very nice people running the service and the food is always good. stars!
40%	i love this place! very nice people running the service and food is always good. thanks!
60%	i love this place! very nice people running the service and food is always good. thanks!
80%	i love this place! good service and fast. exceptional customer service!
W2V-STEM	
20%	i love this place ! very nice people running the cafe also the service is always great . stars !
40%	i love this place ! very nice people running the cafe also the service is always great . stars !
60%	i love staff gym ! extremely knowledgable people running unbelievable boutique also unbelievable service was always great . stars !
80%	i love staff gym ! extremely knowledgable people running unbelievable boutique also unbelievable service was always great . stars !
GWN-STEM / NWN-STEM	
20% (0.075 MIN_{sim})	i quality this quantities ! very nice people running the cafe and the quantities is always good . stars !
40% (0 MIN_{sim})	i quality this quantities ! very nice people running the movie and the quantities is always good . movie !

Table 14: Example outputs for a Yelp evaluation line with noun RE

Yelp - Verb *RE*	
Input:	my favorite spot for lunch . they are quick and the food is always consistently excellent .
RE:	ordered
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	my favorite spot for lunch . i ordered quick and the food is always consistently excellent .
40%	my favorite spot for lunch . i ordered quick service and the food is always consistently excellent .
60%	our favorite spot for lunch . i ordered quick and always fresh and excellent .
80%	fun spot for lunch . i ordered sashimi it was excellent .
SMERTI-RNN	
20%	my favorite spot for lunch . i ordered quick and the food is always consistently excellent .
40%	my favorite spot for lunch . always ordered quick delivery food and always consistently excellent .
60%	my favorite spot for lunch . always ordered quick and always always excellent .
80%	great spot for lunch . everything we ordered was excellent .
W2V-STEM	
20%	my favorite spot for lunch . they are quick and the food ordered always teriyaki excellent .
40%	my favorite spot for lunch . they are quick raw the lamb ordered always teriyaki excellent .
60%	my favorite spot got lunch . they were takeout raw pulled lamb ordered picked teriyaki excellent .
80%	my favorite spot got lunch . they were takeout raw pulled lamb ordered picked teriyaki excellent .
GWN-STEM	
20% (0.1 MIN_{sim})	my favorite spot for lunch . they are quick and the food is always consistently excellent .
40% (0 MIN_{sim})	my favorite spot for lunch . they dropped quick and the food dropped always consistently excellent .

Table 15: Example outputs for a Yelp evaluation line with verb RE

Yelp - Adjective *RE*	
Input:	bad ass vegan options ! everything we had was awesome ! keep it rad guys thanks !
RE:	mexican
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	bad ass mexican options ! everything we had was awesome ! keep it rad guys thanks !
40%	bad service and mexican options ! everything we had was awesome ! keep up your rad guys thanks !
60%	great mexican food ! we had was awesome ! keep it up thanks !
80%	best mexican food ! wings are awesome ! love the service thanks !
SMERTI-RNN	
20%	bad ass mexican options ! everything we had was awesome ! keep it rad guys thanks !
40%	bad little mexican options ! everything we had was awesome ! keep to crowded and funny guys thanks !
60%	best mexican food ! we had was awesome ! keep coming back thanks !
80%	best mexican cuisine ! super awesome ! great service and wonderful thanks !
W2V-STEM	
20%	bad ass mexican court ! everything we had was awesome ! keep it rad guys thanks !
40%	bad ass mexican court ! chinese we had was awesome ! keep it rad guys thanks !
60%	bad ass mexican court ! chinese we experienced was outstanding ! keep it hall guys thanks !
80%	shitty ass mexican court ! chinese visiting experienced is outstanding ! keep just hall folks thanks !

Table 16: Example outputs for a Yelp evaluation line with adjective RE

Yelp - Phrase *RE*	
Input:	cheap ! employees are nice ! great place to go to have a tuna sandwich !
RE:	come back
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	cheap ! employees were nice ! great place to come back and have a tuna sandwich !
40%	cheap ! employees were nice ! great place to come back for a tuna shop !
60%	cheap ! employees are rude ! ill come back for a tuna sandwich !
80%	cheap ! employees are rude ! ill come back for a short time !
SMERTI-RNN	
20%	cheap ! employees are nice ! great will come back and have a tuna sandwich !
40%	cheap ! employees are nice ! great will come back for a tuna meal !
60%	cheap ! employees are amazing ! will never come back with a tuna and cheese !
80%	cheap ! employees are amazing ! will definitely come back with a friend !
W2V-STEM	
20%	come back employees are nice great place to go to have a tuna sandwich
40%	come back employees are nice great place to go to have a tuna sandwich
60%	come back employees are nice great place to go to have a tuna sandwich
80%	come back didn't even patio our server patio would be our server tuna sandwich

Table 17: Example outputs for a Yelp evaluation line with phrase RE

News Headlines - Noun *RE*	
Input:	trump's tactic twitter fights distract from absence of policy positions
RE:	video
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	trump's election campaign video exposes hypocrisy from absence of policy positions
40%	trump's campaign says twitter video is from a surge of policy positions
60%	trump's latest video captures news from the power of controversial positions
80%	trump's new video of
SMERTI-RNN	
20%	trump's america's twitter video goes from absence of policy positions
40%	trump's instagram twitter video goes from a state of policy positions
60%	trump's instagram video looks from women of russia positions
80%	trump's pick video about thursday
W2V-STEM	
20%	trump's tactic twitter fights distract from absence of policy video
40%	trump's photo twitter photo photo from photo of policy video
60%	trump's photo instagram photo photo from photo of election video
80%	trump's photo instagram photo photo during photo captures election video
GWN-STEM / NWN-STEM	
20% (0.075 MIN_{sim})	evidence plan twitter battle distract from lack of plan location
40% (0 MIN_{sim})	evidence plan twitter battle distract from lack of plan location

Table 18: Example outputs for a news headlines evaluation line with noun RE

News Headlines - Verb *RE*	
Input:	identity theft it's not just for grownups !
RE:	save
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	identity theft it's not to save grownups !
40%	identity theft could save grownups !
60%	identity theft could save grownups !
80%	let's save money !
SMERTI-RNN	
20%	identity theft it's not to save grownups !
40%	identity theft doesn't save grownups !
60%	identity theft doesn't save grownups !
80%	ways to save yourself !
W2V-STEM	
20%	identity theft itsave not just for improve !
40%	improve cell itsave not just for improve !
60%	improve cell itsave not just for improve !
80%	improve cell itsave not soon improve improve !
GWN-STEM	
20% (0.1 MIN_{sim})	space theft it's not just for grownups !
40% (0 MIN_{sim})	space space space's not just for jeopardy !

Table 19: Example outputs for a news headlines evaluation line with verb RE

News Headlines - Adjective *RE*	
Input:	hasan minhaj comparing donald trump to stereotypical indian uncles is gold
RE:	american
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	the american comparing donald trump to create indian uncles is gold
40%	the american comparing donald trump to artists who made me uncles
60%	the american evolution of trump to win is gold
80%	defining american anti trump struggle
SMERTI-RNN	
20%	this american comparing donald trump to the indian uncles is gold
40%	the american comparing donald trump to tell uncles is gold
60%	african american university president trump to congress is gold
80%	african american leader on trump rally
W2V-STEM	
20%	hasan minhaj comparing donald trump to stereotypical american uncles is gold
40%	hypochondriacs hypochondriacs comparing donald trump to hypochondriacs american uncles is gold
60%	hypochondriacs hypochondriacs comparing donald trump to hypochondriacs american uncles is gold
80%	hypochondriacs hypochondriacs discuss donald trump but hypochondriacs american hypochondriacs is soccer

Table 20: Example outputs for a news headlines evaluation line with adjective RE

News Headlines - Phrase *RE*	
Input:	kindergarten teacher allegedly drinks beer at school
RE:	donald trump
MRT/RRT	**Generated Output**
SMERTI-Transformer	
20%	donald trump allegedly drinks beer at mexico
40%	donald trump allegedly drinks beer president
60%	donald trump makes beer cans
80%	donald trump will not apologize
SMERTI-RNN	
20%	donald trump allegedly drinks beer at now
40%	donald trump allegedly drinks beer in big race
60%	donald trump takes a beer debate
80%	donald trump cabinet challenge
W2V-STEM	
20%	kindergarten teacher donald trump drinks beer at school
40%	kindergarten teacher donald trump drinks beer at school
60%	kindergarten teacher donald trump drinks beer bigger
80%	kindergarten teacher donald trump drinks beer bigger

Table 21: Example outputs for a news headlines evaluation line with phrase RE